MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS

AD-A168 356

RSRE
MEMORANDUM No. 3827

ROYAL SIGNALS & RADAR ESTABLISHMENT

THE REPLAY TECHNIQUE: THE CONCEPT. INITIAL
EXPERIENCE AND PROPOSED DEVELOPMENTS

Authors: A Jackson and G J Onslow

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
RSRE MALVERN,
WORCS.
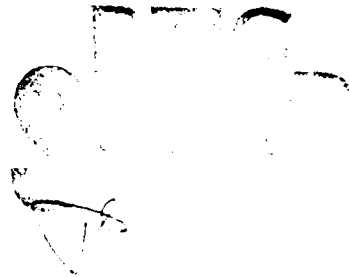
RSRE MEMORANDUM No. 3827

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 3827

TITLE: THE REPLAY TECHNIQUE: THE CONCEPT, INITIAL EXPERIENCE AND PROPOSED DEVELOPMENTS

AUTHORS: Alistair Jackson and George J. Onslow

DATE: November 1985

SUMMARY

A system is described for capturing activity at the Human-Computer Interface in complex systems. This system, implemented in a Simulated Air Traffic Control Environment, allows for the subsequent replay of all task related, audio and visual inputs to the controller, synchronised with a record of his own manual and spoken system inputs. Potential applications for such a facility are discussed and three pilot studies considering the tool's usefulness in examining operator workload and assisting in system development are reported. A revised version of the facility, for use in an enhanced simulation environment is also described.

CARDPB Ref: 2.7.1.1

RSRE Task   69℃

CAA Sponsor   Ch.Sc./CAA

CAA Liaison Officers: C L Hynam, CR(8)
                        M J Asbury, RD1

RSRE MEMORANDUM 3827

THE REPLAY TECHNIQUE: THE CONCEPT, INITIAL EXPERIENCE AND PROPOSED
DEVELOPMENTS.

Alistair Jackson and George J Onslow

LIST OF CONTENTS

1.0 INTRODUCTION

This memo describes the 'REPLAY' facility, which was implemented on
the AD4 (formerly G3) Air Traffic Control Simulator of the Terminal
Control Systems Development Group (TCSDG) [REF 6,11]. This facility
provides for the recording of much of the information flow at the
man-machine interface and allows for its subsequent 'replay' in real
time, or a stepped presentation in small increments for detailed
analysis. A consequence of this approach is that it provides a
history of the external observable behaviour of the system which can
be of use in the general debugging process as well as in human factors
investigations.

In Section 2 the objectives prompting the initial development of the
Replay concept are outlined. Section 3 provides details of an initial
implementation, and Section 4 describes the procedures adopted in
three brief pilot studies and discusses their results. In Section 5,
an expanded version of the original facility, intended for
implementation on the new AD4 VAX 11-780 simulator system, is
discussed. Finally, Section 6 provides a brief summary and
conclusions.

2.0 BACKGROUND TO THE DEVELOPMENT OF REPLAY

The original impetus for the development of the Replay concept derived
from three interests- the study of the Air Traffic Controller's
'Picture', measures of workload and their validation, and the
development of tools and methodologies for data collection at the
human-computer interface. These interests converged at a time when
trials on the utility of ground-based height measurements provided an
opportunity for the collection of data if a suitable implementation
could be quickly and cheaply provided. The following paragraphs
describe the common objectives which derived from the three areas of
interest.

2.1 Justification for the development of Replay

**2.1.1 The Study of the Air Traffic Controller's Picture as a Mental
Model.**
AD4 has been interested in study of the Air Traffic Controller's
(ATCO's) 'picture' for some time [REF 19]. An appreciation of the
potential significance of the concept was first gained during work on
Interactive Conflict Resolution (ICR) [REF 20], a tool which allowed
ATCOs to examine computer-generated extrapolations of air traffic
situations as a planning aid. 'Picture' is a term frequently employed
by ATCOs to refer to an aspect of their internal representations of
the traffic situation; in other contexts this would be called their
'mental model'. At changeover between controllers the incoming ATCO
may spend up to 10 minutes watching over the shoulder of his
predecessor 'building his picture' before taking over. A number of
aspects of picture are discussed in other papers [REFS 19,21]. The
second of these papers discusses the use of replay and deals with some
of the issues covered in later sections of this memo, placing more

emphasis on the psychological aspects. One of the most important properties claimed for picture is its dynamic nature and controllers place some emphasis on maintaining its currency. It was thus considered important when some controllers expressed doubts, during the ICR trials, as to whether reliance on some forms of automatic assistance would interfere with their ability to retain the picture.

The preliminary work on picture led to the recognition that there was an acute shortage of techniques suitable for collecting useful data. One possibility, a method which had proved useful in other investigations of user's models, and had been employed with some success in ATC at Manchester Control Centre [REF 21,19] was the collection and analysis of verbal protocols. A verbal protocol is a running verbal commentary generated by an individual performing a task. The commentary describes the processing and the operations which the user thinks he is performing. While verbal protocols have been employed in a wide variety of contexts [REF 2,17] there is theoretical controversy concerning the range of their usefulness [REFS 4,5,10,18 all provide interesting discussion]. In spite of this controversy there is some expectation that protocols appropriately analysed might at least provide insight into some of the processes which controllers actually employ, and the objects to which those processes are applied. In live ATC situations however there is a further major difficulty. ATCOs make such substantial use of the verbal and auditory communication channels that the generation of concurrent verbal protocols is impossible without risking unacceptable interference with task performance. This means that the only way in which a protocol can be obtained is retrospectively, after the task has been completed.

Retrospective protocols are not only more contentious from a methodological viewpoint, but there are problems in their collection, principally arising from the continuous nature of the ATC task. Generally retrospective protocols can be collected in two situations:

a)   Where tasks are of a very short duration.

b)   Where the task can be halted unexpectedly and the subject asked to generate a protocol.

Although a variation on the second approach has been used in ATC research [REF 3,15] it essentially functions as a probing method to investigate points of interest. It can only be operated in simulation situations since the real task cannot be interrupted and it can only capture, in any detail, a short period prior to the interruption. As indicated, the establishment of picture takes between 5 and 10 minutes and there is further anecdotal evidence which suggests that the time span of a fully developed picture is around 20 minutes. This 20 minute period corresponds to the time which an aircraft typically requires to traverse a sector and thus represents the average time for a complete replacement of all aircraft in the sector. After such a period the controller might logically be expected to have complete information available.

Two related ideas led to the notion of replay as an approach. The first is that an individual faced with the input of information,

interprets that information in terms of a repertoire of stereotypical
structures or schemas which are cued by the person's general context
state, determined in turn by the external environment and the internal
state of the individual and his preceding processing. If an
individual is faced with a novel situation, he will attempt to deal
with it by reference to prior experience. He will invoke those
schemas which seem to be most similar and appropriate to his current
situation. If he could be faced with an identical situation, we would
expect the same schemas to be invoked as on the prior occasion.
Actually, an individual can never truly be faced with exactly the same
situation twice since he brings a different internal state to the
event by dint of his prior experience with the situation.
Nevertheless, we might expect that the more accurately we can
re-create the external conditions of a situation, the closer the
similarity will be between procedures and processes invoked in the
individual. The objective of replay is to capture these external
conditions as accurately as possible.

The schema notion is very closely linked to the second idea, deriving
from a more classical memory research tradition, that of state
specific memory. This suggests that retrieval of information,
especially recognition memory, is facilitated when the individual
re-experiences the same physical situation. In essence the schema
approach is almost an explanation of this classical observation.
While we would hold that concurrent verbal protocols at best offer
insight into underlying knowledge and processes and that careful
interpretation and analysis is required to make appropriate
inferences, the situation with retrospective verbal protocols becomes
more complex. There is a very real danger that the subject will
produce a post-event interpretation of what actually took place edited
in the light of the subject's goals, his knowledge of the immediate
consequences of his actions, and his interpretation of the
experimenter's objectives. Effectively he will actively reconstruct
events from long term memory. There has long been a distinction
between the processes of recall and recognition in descriptions of
long-term memory retrieval. Recall is the active, reconstructive,
retrieval, 'trying to remember' to which we have already alluded; it
is limited in sensitivity by the reason for remembering. Recognition
on the other hand, is much more sensitive and, subjectively, appears
to function in a less controlled manner, the correct stimulus evoking
a memory of an event which we were not aware we could remember. The
parallel with the schema approach is clear. If we can provide the
correct prompts we might be able to obtain a clearer access to the
knowledge and processes that interest us.

2.1.2 Measurement of the Air Traffic Controller's Workload
The second impetus for the development of the Replay concept derived
from an interest in measurement of the ATCO's workload and means for
the validation of such measurements. Such measurement is important in
the planning and evaluation of modifications to airspace and sectors,
and in assessing the effectiveness of revisions in control procedures.
A number of workload measures have been employed, involving varying
degrees of interference with the task and amounts of discomfort for
the controllers, but one of the most acceptable has been the Observer
Method, evolved by the Directorate of Operational Research and

3

Analysis (DORA, now DR) within the Civil Aviation Authority [REF 1, 16]. In this method an ATCO, familiar with the sector, watches 'over the shoulder' of the duty controller and rates workload on a numerical scale at fixed time intervals. This method works well where the aim is to establish some statistic describing the typical workload associated with particular configurations, but it fails to capture the range of 'experienced workloads' for different controllers with different styles. In addition there are three areas of uncertainty associated with the observer method.

1. The observer must translate from his own viewpoint to how he thinks the controller sees it.

2. The rating scales are very short being only 3 or 4 points depending on the version employed. Further, the category descriptions of the rating scales have not been subjected to validation checks.

3. There is no ready way to evaluate the effects arising from using such a limited number of observers.

The Replay approach seemed to offer a way round some of these difficulties and uncertainties. Firstly, if controllers could provide a workload rating analysis on their own task performance the difficulties caused by difference in viewpoints could be minimised. While task demands preclude rating during the control task, ratings could be generated during a replay of the situation. Secondly, if the ratings generated by controllers viewing a replay closely corresponded to those produced by concurrent, external observers using the conventional rating method, the validity and the value of the method could be appraised independently. Thirdly, the replay could be presented to a large sample of observers giving an insight into any effects which might arise from using a small number of observers.

2.1.3 Replay as a Tool for System Development.
The final spur for the production of the Replay technique was the recognition that it might become a useful tool in the development and debugging of systems, especially when an iterative prototyping approach had been adopted for system development. This approach is typical of much of the work carried out by AD4, and indeed is becoming widely recognised, under the title of 'Very Rapid Prototyping' as a necessary technique where complex, man-machine systems are required. [REF 8,13]. When a system is being developed in this way, with constant trials, revisions and the addition of further modules to expand its capability, two problems can assume increased importance: debugging; and maintaining an adequate description of the system.

Clearly, a system under constant revision, is constantly subject to new bugs, even with the most careful software production. Debugging becomes a more routine and acceptable process but is something which has to be carried out more often. Tools and strategies are developed to cope with the problem; examples include better editors, automatic version numbering and the use of history files which log the activities of a system during a run. Replay represents an extension to the history file. In complex man-machine systems it is often

helpful to know not only the activity of the system prior to a fault developing but what the operator was doing, both in terms of physical interaction with the system and in terms of the procedures and processes in which he was engaged. This can be very helpful in identifying areas of difficulty in the function of the total system. If Replay can provide a record of activity at the interface it can be used in conjunction with the history files to assist this type of debugging. Usefulness will almost certainly reflect the completeness of both types of record.

Maintaining an adequate description of a system under continuous development is an exceedingly difficult problem. The overhead in keeping documentation current is high and 'snapshot' copies of the system, carefully annotated, have to be taken at suitable intervals. Replay can make a contribution to the quality of the snapshot by capturing something of the nature of the interaction with a particular system.

2.2 Objectives of the Replay System

These themes converged to produce the idea of replay as a system with a number of features.

1. It had to capture all the information inputs available to a controller during his run on the simulator. Essentially this included the radar picture, the R/T input and any additional auditory input from the environment, and flight strip displays or their equivalents.

2. It had to be able to re-present the information, in the correct sequence, in real time or at a variable controlled rate.

3. The recording system had to be completely non-intrusive with respect to the control process.

4. The development time of the trial system had to be very brief, and the system had to place minimal demands on effort and resources.

3.0 DESCRIPTION OF THE INITIAL IMPLEMENTATION

3.1 General Description of the Replay System.

As indicated the main objective of the facility was to capture activity across the man-machine interface. During the period in which the ideas about the facility were developing a comparatively small scale simulation was being conducted to consider the potential usefulness of fixed position, ground based height measurement equipment for the reduction of vertical separations above Flight Level 290 [REF 14]. This simulation involved a single controller and a blip-driver, who interacted with the simulator to produce aircraft manouevres in response to the controller's ATC instructions. A description of the simulator and the technical details of the replay implementation are provided in the main body of this section.

The controller was seated in front of a Plan View Radar Display (PVD)
with a computer printout which replaced the flight strip boards
normally employed in ATC tasks. The controller was free to annotate
the listings in much the same way as on conventional strips.
Communication with the blip-driver was by means of standard R/T
headsets with boom microphones and a hand operated transmit switch.
Sources of input to the controller were therefore limited to the PVD
and the flight strip listings in the visual modality, and to the R/T
link plus any external ambient noise and verbal asides (including
talking to themselves) in the auditory modality. The initial
implementation aimed at capturing these four sources. In the
recording mode (Figure 1) the auditory material was captured on one
track of a conventional reel-to-reel magnetic tape recorder and the
details required to re-draw the radar picture were stored digitally on
the magnetic disc storage peripheral to the computer driving the
simulation. The audio recorder eavesdropped on the controller's
headset and microphone via the R/T, telephone system. The annotated
flight strip listings were retained. Synchronisation between the
radar picture recording and the auditory data was achieved by
recording a tone on the second track of the tape recorder at the
beginning of each radar update. In the replay mode, the tape recorder
drives the entire system (Figure 2). A computer program detects the
tones on the second track, retrieves the appropriate radar picture
from the discfile and writes it to the plan view display in real time.
The replay can be halted at any time simply by using the pause control
on the tape recorder.

A more detailed description of the simulator follows.

3.2 Description of the ATC Simulator and the Interaction with Replay.

The work on Replay was based on an available air traffic simulator
that had been used for another experiment [REF 14]. This simulator
was a simplified version of one used by the Terminal Control Systems
Development Group for their experiments [REF 6]. It was written in
Coral 66 and ran on a multiple processor, Computer Technology Limited
(CTL) Modular-One computer system.

The simulator modelled the airspace in Scotland but used hypothetical
air routes (loosely based on actual routes). A map of the routes is
shown in Figure 3. Up to one hundred and forty aircraft could be
simulated each one being 'flown' either automatically or manually by a
'blip' driver. Each 'blip' driver was equipped with VDU and keyboard
that allowed him to control up to twenty-five aircraft. The simulator
allowed for up to four radar screens for ATCO's and up to four 'blip'
driver positions. The simulator consisted of a number of tasks that
ran concurrently. An aircraft's position was updated once every ten
seconds (the time for a simulated radar revolution). A global data
base was maintained that held positional and other information for
each aircraft in the system.

The simulator consisted of 5 concurrent processes, communication and
synchronisation between processes being message based. Figure 4
illustrates the software structure and the processes are described
below :-

FPC (Flight Plan Control)
This process dealt with the activation of new aircraft at the correct time and the removal of aircraft that have left the system.

NAV (Navigation)
This process controlled the navigation of aircraft in the system. Flight path demands (changes in flight-level, heading etc.) were sent to APU.

APU (Aircraft Position Update)
This process performed the aircraft flight path position updates in response to a stimulus from NAV. Each aircraft in the system was updated every ten seconds (the radar revolution time). After all the aircraft had been updated a stimulus was sent to the DIS process and the global data base was updated as appropriate.

COM (Communications Interface)
This process provided the communication between the 'blip' drivers and the simulator. Commands were received from the 'blip' drivers' keyboards and appropriate messages sent to NAV. Reports from NAV were displayed on the 'blip' drivers' displays. COM also 'flew' the aircraft that were not under a 'blip' drivers control but being controlled automatically.

DIS (Display)
The DIS process drove the radar displays that showed plots of airborne aircraft within the area of the display, superimposed upon a video map that showed airports and radio beacons. Each aircraft plot had associated with it a two line plaque that gave the aircraft callsign on one line and its flight-level and destination on the other. There was also a heading vector that joined the plot to the predicted plot position in one minute's time. The radius and centre of each radar display were set at run-time. Up to four displays could be used.

In addition all the tasks wrote to a 'History' file whenever a significant event occurred. At the end of a simulator run this file thus contained a full history of that run.

A fuller description of the TCSDG simulator appears in [REF 6].

Also available was a traffic sample generator. This generated traffic samples for each route to a given loading. Types of aircraft, callsigns, flight-level etc were also automatically generated.

It had been decided at the outset that adding Replay to the simulator should ideally have not involved any changes to the simulator. This requirement was met. As it was required to replay both the radar picture and the ATCO/Pilot dialogue it was decided to use an available UHER two track audio tape recorder. One track was used to record the ATCO/Pilot dialogue, the other was used to record synchronisation tones for the replay. This tone appeared at the beginning of every radar rev, being placed there by the recording program. The replay program used this tone to maintain synchronisation between the replayed picture and the replayed dialogue. The tone was not encoded

in any way and thus just provided a 'next rev' type of
synchronisation. Synchronisation was established at the beginning of
a Replay run and the tone enabled it to be maintained.

Adding the Replay facility to the simulator consisted of one piece of
hardware, the Tone Generator/Detector, and two programs, one to record
relevant data during a simulator run and another to replay that data.

The Tone Generator/Detector interfaced to the Modular-One computer and
had two modes of working. The first was used when the original
simulator run was taking place. The recording task caused the Tone
Generator/Detector to output a short (0.25 secs) tone to the tape
recorder at the beginning of each radar rev. The second mode was used
during replay. The Tone Generator/Detector detected the tone that was
output from the tape recorder and generated an interrupt to the Replay
program.

The recording program ran as an additional task in the simulator
system. Once per radar rev it read the common data base and appended
the required data (aircraft position, height etc.) for each airborne
aircraft in the system to a file and caused the Tone
Generator/Detector to output a short tone to the tape-recorder.

The Replay program displayed a radar picture identical to that
generated by the simulator onto a radar display. The radius and
centre of the display was set at run time (and could be the same or
different from that set for the original simulator run). The program
when interrupted by the Tone Generator/Detector read the aircraft data
for the next radar rev from the data file dumped by the recording
program. Using this data the aircraft plots and plaques were
re-created for that radar rev. The Replay program also had facilities
that allowed the user to re-establish synchronisation in the event of
it being lost.

4.0 PILOT STUDIES AND THEIR OBJECTIVES.

Pilot work using the facility was partitioned into two studies and a
brief examination of the use of video to extend the data capture. The
first examined the ability of two controllers to produce workload
ratings on replay and was carried out in the closing stages of the
Height Measurement Evaluation. The second study, which had to be
delayed until an increased number of traffic samples were available,
examined the production of fuller verbal protocols under immediate and
delayed replay conditions.

4.1 Pilot Study 1: Workload Rating

4.1.1 Description
Initially there were hopes that the Height measurement trials would
provide an opportunity to establish whether the replay approach
offered a feasible means of examining both the workload rating
procedures and the production of verbal protocols describing control
process activity. Unfortunately, the need for several revisions of
the traffic pattern and the route structures during the height
measurement simulation meant that there was difficulty in collecting

8

more than one sample for each of the two ATCOs available.

The single sample obtained was used to establish whether the two available controllers found re-viewing under replay acceptable and, since the production of fuller verbal protocols would have introduced a second completely novel factor, workload ratings only were obtained. The sector used was imaginary in terms of traffic composition and route structure but was based on the geography of the West of Scotland. Traffic level was 50 aircraft per hour with a wind of 60 knots from 110 degrees.

Data was collected under four conditions:

> i) Recording of control session with Controller A, no rating observer present.

> ii) Recording of control session with Controller B, observer workload ratings by Controller A.

> iii) Replay of session by Controller A (i), observer ratings by both A and B.

> iv) Replay of session by Controller B (ii), observer ratings by both A and B.

No observer data was collected during Condition (i) since it was undesirable that Controller B should take active control of a traffic sample which he had already viewed as an observer.

The observer rating procedure in conditions ii, iii, and iv was essentially similar to that used by DORA. The observers, familiar with the sector through prior experience with other samples, rated the workload, integrated over two minute periods, on a five point scale (1=very low, 2=low, 3=average, 4=high, 5=overloaded). In condition (ii) the observer sat slightly behind, and to one side of the controller, in a position where he could observe the radar display and listen in on the controller's transactions. During the replay ratings (conditions (iii) and (iv)) both observers sat in front of the radar and listened to the audiotape, with no communication between them.

4.1.2 Results

The ratings obtained under the various conditions are shown in Figures 5(a) and 5(b).

> 1. The controllers were quite happy to rate their own replays, and although one described it as "a levelling experience" both showed keen interest in observing their own handling of the traffic.

> 2. On the data for Controller B a crude comparison of the ratings provided by the controller himself in replay and the observer was possible. A crude correlation test, run between the two data sets, yielded a coefficient of 0.71, (significant at the $p < 0.01$ level). Although it is very difficult to generalize from such limited data, especially when the results are

9

constrained by a five point rating scale the apparent
consistency between the two rating sets was encouraging.

3. The insistence, by both controllers, on using a "4 to 5" rating
    value, (entered as 4.5 in the correlation calculations),
    suggested that the rating scale required at least one more
    category corresponding to "very high" traffic.

4. According to controller comments, visual scanning of the rather
    complex sector represented a substantial component of the
    workload at higher levels.

4.2 Pilot Study 2: Verbal Protocols

4.2.1 Description

The second study involved the collection of verbal protocols during
the replay phase. The same simulation of a fictitious Scottish sector
was employed with new traffic samples of the same densities as in
Pilot Study 1. A number of trial runs were necessary before a
procedure for obtaining adequate protocols was established. It
quickly became apparent that a controller had to become familiar with
the process of producing such protocols as well as with the sector.
Initial protocols were staccato and terse but, after a number of
sessions, fuller, more fluent protocols were produced. Once
relatively full protocols were available a slightly more organised
study was conducted using one of the controllers from the original
study and one new controller.

The objectives of the study were:

i) To collect original control transcriptions and related
    retrospective protocols as data on which to establish means of
    qualitative or quantitative analysis and with which to gain
    insight into control processes.

ii) To produce protocols under immediate and delayed replay
    conditions in an effort to establish whether qualitative
    differences could be discovered deriving from the effects of
    memory and knowledge of outcomes in the immediate replay.

Two samples were prepared and four protocols were generated; each
controller running both samples and generating one immediate protocol
and one delayed protocol. The immediate and delayed protocols were
generated on different samples for the two controllers. Immediate
protocols were obtained by running the replay system as soon as the 1
hour simulation run was completed so that description occurred 1 hour
plus a few minutes after each individual control event. In the
delayed condition the replay procedure was undertaken some 10 weeks
after the original recording. In both cases the protocols were
produced by allowing the controller to sit on his own, wearing his
headset, viewing the replay and hearing the transactions over his
earphones. His protocol was recorded on the headset microphone on a
second audio recorder. The second track of this audio recorder
received the synchronization pulses from the replay recorder and thus
the protocol recording could not only be synchronised with the initial

control protocol for analysis but could be used to drive the simulator to produce a secondary replay providing the radar picture and the controller's description of his control. Unfortunately, the synchronisation pulses were lost on one of the delayed control conditions.

4.2.2 Results
The volume of data generated in this study was intimidating, eight hours of protocol (four original + four replay) had to be transcribed. As a consequence the results and conclusions are based on two phases of analysis. Phase 1 results have already been reported [REF 21] and shall be summarised only briefly. Phase 2 continues as techniques develop [REF 5] and effort becomes available.

Phase 1
Several observations and conclusions were reached as a result of this pilot study and the preliminary protocol collection runs. These are summarised as follows:

   i) As already indicated, the controllers require a period of familiarisation with protocol production before an extensive protocol could be produced. Even then the fluency of the protocols tended to suggest that the controllers were aiming their utterances at the observer rather than introspecting on the processing associated with the task. Leplat and Hoc [REF 10] discuss the tendency for subjects to modify their verbalisations to comply with their representation of the observer's understanding of the skill. Both controllers in the present exercise were experienced as instructors and this may well have significantly affected the style and level of their verbalisations. It was this factor which suggested that leaving the controller with no audience while generating the protocol and encouraging him to 'talk to himself' might prove most fruitful.

  ii) There were indications that attending to the R/T exerted an inhibitory effect on the generation of the protocols. The R/T was not under the speaker's control and could interrupt his verbalisations at any stage. While the replay could have been placed under the speaker's control this would have interfered with the uniformity of the time delays in the immediate replay conditions. In this context, there is also a possibility that one's own voice on tape is a more potent distractor than another speaker's, although this has not been verified.

 iii) The protocols also served as a prompt for discussion with the controller and, in fact, one of the procedures adopted was for the analyst to study the replay protocol and produce a description of what he thought was going on, then to step through the protocol with the controller discussing the validity of the description. Figure 6 shows a sample of the output of this procedure. The discussion also embraced the use of flight strips and a number of observations and findings by previous researchers were confirmed including the primary importance of the manually updated flight strips during planning phases for looking ahead, and the greater relative use

11

of radar during the implementation and supervision of plans.
This is also supportive of the findings described in the first
part of the IFIP paper [REF 21] which emphasizes the importance
of flight strips in the initial establishment of picture.
There were controller differences in the way in which flight
strips were employed but the general strategy was very similar
to the 'flight level strategy' described by Leplat and Bisseret
[REF 9]. The authors consider that this general use of the
replay as a discussion prompt also shows great promise as an
aid in system, and procedural development in iterative
prototyping environments. An interesting account of IBM's use
of similar tools is provided in a paper by Neal and Simons [REF
12].

Deeper analysis of the protocols has proved to be very difficult and
demanding. A particular difficulty is caused by the number of
pronouns used in the replay protocols, referring back to either the
replayed audio track or the radar picture. While transcripts of the
two protocols (control and replay) can be laid out together, it
becomes necessary for the analyst to be able to view the radar image
during the analysis process. Normally provision of this facility
would present little problem but the Divisional Computer Replacement
Programme has meant that it was not available for substantial periods
of time. It is strongly recommended that such a facility be provided
in comparable analytic situations.

4.3 Use of a Video Recording Facility to Study the Controller's Direction of Gaze.

In addition to the pilot studies, brief trials were conducted to
examine the usefulness of a video recording facility to augment the
information available. The time synchronization available with replay
can tell the analyst that the controller made a particular remark or
decision while a certain piece of information was available on a
display, but it does not tell him whether the controller was looking,
or had just looked at that display. Eye movement recording would
indicate the controllers viewpoint within one or two degrees of visual
angle but represents a very substantial overhead in analysis and may
be both restrictive and uncomfortable for the controller. An
alternative, and very much cheaper, approach involves the use of video
recording of the user's head and eyes to identify the direction of
gaze at any instant. These trials were conducted to discover whether
such recordings would provide sufficient resolution to establish which
display or part of a display a controller was fixating, and to see if
an analyst could make a reasonable categorisation on a single real
time run of the tape. It should be noted that knowing that someone is
looking at something does not prove that they are seeing or processing
that item, but it is supportive evidence.

The recording technique was essentially as before except that a video
recorder with twin audio channels was substituted for the audio
recorder. The two channels were used for the audio track and the
timing pulses respectively, and replay could take place exactly as
before. Recordings were made for two controllers on the different

traffic samples. A simple analysis was carried out using an interactive computer programme which allowed the analyst to press one of a number of keys, corresponding to different categories of activity. The program computed the proportions of time spent on each category. These initial trials used very dated video equipment with corresponding quality in the images produced; however it clearly demonstrated the feasibility of such analysis and similar techniques have since been employed successfully during other simulated air traffic evaluations. The analysis performed revealed, that in the configuration employed, controllers spent close to 50% of their time on each of the two information sources, PVD and flight strip analogues. (Controller A, 48% PVD, 52% FS; Controller B 52% PVD, 48% FS). Repeated categorisations of the direction of gaze appeared to produce very similar results although there was no way in which this could be verified statistically with such a small sample.

5.0 EXPANSION OF IMPLEMENTATION TO AD4, VAX 11/780.

It is intended to re-implement Replay on the new TCSDG simulator that runs on a Vax 11/780. This new simulator is basically the same as the original but rewritten for the new machine. No major problems are foreseen in this transfer, the only possible problem area being the manufacture and interfacing of a new Tone Generator/Detector. Ideally the new Tone Generator/Detector will use some form of encoded pulse to enable synchronisation to take place automatically. No large design changes are envisaged for the new version of Replay but some changes will occur due to differences between the Modular-One and Vax and also due to the different dialects of Coral 66 used by the two machines. A full description of the Vax based simulator is available [REF 7].

The TCSDG group also make use of a computer based Air Traffic Management (ATM) system that runs in conjunction with the simulator. It is beyond the scope of this document to discuss the ATM system in detail but basically it provides displays and other aids to generate a more strategic approach to Air Traffic Control. The ATM generates a management plan for each aircraft, leaving short term tactical control to the human controller. A fuller description of the ATM system is also available [REF 11].

It is hoped to extend the replay programs to replay the ATM system. This will be a more difficult task than replaying the simulator for two reasons:-

1. The ATM system has a number of displays that update at random times. The simple synchronising pulse technique used for the simulator replay will not cope with these as to 'snapshot' the database every time a display update was made would produce an unacceptable amount of data. A number of techniques are possible for the synchronisation but some experimentation will be necessary before deciding on the most appropriate one. A possibility is to maintain basic synchronisation using the radar rev as before but to use the various timing mechanisms provided by VMS (the operating system of the VAX) to achieve

13

precise synchronisation within the radar rev. Data would thus only need to be dumped once per radar rev.

2. The ATM system is far more complicated than the simulator and production of software to generate displays identical to those within the ATM system but using the recorded data would be a major undertaking. This means that the original ATM display processes would be used to generate the displays but that the stimuli to those processes would be generated by specially written Replay software. The ATM system was not written with the addition of replay in mind and thus some of the processes that would be replaced by the Replay stimuli generating processes may not be easily detached from the rest of the ATM system. Had the ATM system been designed from the outset to facilitate the addition of Replay then no such problem should have arisen.

5.1 A General Recording Facility.

This section describes an extended recording facility, based on Replay which incorporates a number of data sources in a synchronised record and can be collected semi-automatically during any run of a suitably equipped simulation facility. It is hoped that such a facility will be incorporated within AD4's simulation effort.

5.1.1 Criteria for the Facility.

These criteria represent a refinement and expansion of those for the original Replay implementation in the light of experience from the pilot studies, examination of other techniques and methodologies, and additional thinking on the type of data required particularly for the 'debugging' approach to system development.

1. The non-intrusive nature of the initial data recording process is stressed. Any analysis of the type envisioned is worth nothing if the task has been significantly changed by the data collection process. This places a premium on automatic, or semi-automatic data collection which is not only invisible during operation, but does not require a lengthy and observable set-up period. It also requires that the data collection processes should not place such overheads on the available computing power as will perceptibly modify the performance of the machine components of the system.

2. The potential for data recording should be as complete as possible given the constraints implied by (1) above. The processing load on the system machine can be minimised either using a separate recording processor as in the IBM method [REF 12] or by making significant use of an external processing medium such as the video or audio tape used in the Replay approach. Each has advantages in capturing a form of data difficult for the other. In situations where keyboard or other forms of manual input are very numerous the dedicated filtering processor is almost certainly superior but multiple workstations would either require multiple processors or time-sharing on a single processor with the concommitant

14

prospect of delays in passing the output to the host processor. The tape media solution makes capture and synchronization of speech activity and direction of gaze simpler but capturing keyboard and analogue inputs becomes more difficult. One of the most critical factors is the timing resolution required of the system. In the original replay implementation, since the simulator was radar based, a 10 second interval, the period of the radar update, was a natural choice. However, where keyed inputs are involved measurement resolutions of milliseconds may be a requirement. Clearly capturing display data as often as this would result in an unacceptable memory overhead, so ideally only display changes should be described.

3. The data collected should be capable of analysis at a number of levels or to a variety of depths. Thus we should be able to perform a shallow analysis for a particular class of event or sequence of keypresses or, alternatively, we should be able to integrate and synchronise the information from our different sources to build up a detailed description of any periods of interest. This implies at least two things. The first and more obvious is that we should have adequate synchronisation information to integrate our different sources even if they are not all recorded to the same degree of resolution. Different sources may require different depths of analysis. In particular our experience has indicated that in general the transcription overhead in analysing verbal protocols is very high and the task of deep analysis exceedingly complex; on the other hand the ability to replay the verbal transactions before a particular event of interest and relate them to other information is very useful in understanding what actually took place. Secondly there is the need for tools to help analyse and summarise the substantial quantities of data available. The processing power of the computer should be used to filter and analyse the explosion of material which it makes available. One such tool has already been described in the direction of gaze categoriser where a simple program and single pass provide useful statistics without going to eye movement analysis. The IBM Playback system incorporated similar facilities on the log of keypresses and it does not require a significant leap to conceive of something like a logfile editor which permits the merging of synchronised gaze and keypress files and can search and enumerate instance of user defined, or system detected behaviour sequences.

4. The playback facilities of Replay should be preserved.

5.2 Description of a Possible Revised Replay Facility.

In our revised facility we would wish to capture the verbal interactions with other human operators, the direction of gaze of the controller, the data displayed, and the timing and nature of any keyboard inputs made to the system. The more complex problem of analog inputs to the system such as could be made by a joystick or a rolling ball will be neglected for the moment. Collection of all this material confronts us directly with the differences between the

15

recording methods described in (2) above. A compromise approach
suitable for our ATC radar environment would be to retain an
appropriate radar update for the primary radar displays which could be
dumped to discfile as before, and to record direction of gaze, verbal
transactions and synchronisation information on a video recorder with
stereo audio channels and thus replicate all the facilities of the
original Replay implementation. Keyboard inputs could be logged to
the resolution permitted by the system clock as part of a history file
(Note 1). Changes in tabular displays could be noted in this log and
if required snapshots of their text could also be dumped to disk.
Synchronization between the logfile and the video based record would
be rather more complex than in the previous system. The timing pulses
would retain coarse synchronization between the media but within the
ten second epochs each system would run independently. A millisecond
clock signal can be superimposed on the video image with the timer
started and stopped by the simulator during recording mode. This
clock could be reset by the recording system at periodic intervals to
minimise any small slippage between the media. (Indeed the video
recorder itself could be switched on and off by the recording system
as required to avoid difficulties—if the simulator is set into fast
time mode). During the Replay mode (if it is required) the system
would again be driven by the video initiating the radar updates but
the replay program would be rather more extensive than in the previous
version and would run free between pulses displaying keyboard inputs
and tabular display updates on a separate monitor display, in
accordance with the data stored in the logfile. In addition to the
replay component of the system the other analytic tools described in
(3) would be required.

6.0 SUMMARY AND CONCLUSIONS

1. The concept of Replay has been outlined and the initial
    implementation and pilot studies have been described. It is
    considered that the initial objectives, defined in Section 2.2
    have been met.

2. Pilot Study 1 indicated that the technique would be suitable as
    a means of verifying the validity of the DORA Workload
    Assessment Method should this be considered necessary.

3. Pilot Study 2 indicated that while Verbal Protocols were
    generally informative, the analysis overhead was exceedingly
    high. Further, there is a paucity of fully developed
    techniques for deeper analysis.

4. Video recording proved to be a useful extension of the
    facility.

5. One of the most promising aspects of the Replay tool would
    appear to be its application to the debugging of system
    operation. While deep analysis would be costly over long runs,

Note 1. Such a history file already exists on the AD4 simulation
facility as a general debugging tool.

16

such analysis centered around critical incidents and
interesting system events has great potential. The cost of the
recording itself is very low and uninteresting runs could be
routinely discarded with minimal overhead.

6. A version of the facility, suitable for the revised TCSDG
   System was described.

7. The cost of a replay facility need not be high, particularly if
   the requirements are considered at an early stage in the design
   of the developmemt environment. The authors consider that the
   usefulness of such a facility readily justifies the effort
   required for its provision.

REFERENCES

[1]   Ablett, R.M., *"An Assessment of the Capacity of the Daventry
        Sectors at the London Air Traffic Control Centre October
        1981"*, CAA Paper 82018, November 1982.

[2]   Bainbridge, L., *"Verbal Reports as Evidence of the
        Operator's Knowledge"*, Int. J. Man-Machine Studies, 11,
        411-436, 1979.

[3]   Bisseret, A., *"The Air Traffic Controller's Thinking
        Processes"*. In the Proceedings of the "Symposium on Safety
        and Efficiency: the Next Fifty Years", held to celebrate
        the Golden Jubilee of VNV the Dutch Airline Pilots'
        Association, 1981.

[4]   Ericsson, K.A., & Simon, H.A. *"Verbal Reports as Data"*,
        Psychological Review, 87, p215-251, 1980.

[5]   Ericsson, K.A., & Simon, H.A., *"Protocol Analysis: Verbal
        Reports as Data"*, MIT Press, Cambridge, MA, 1984.

[6]   Finch, W., *"Simulator Software Structure and Facilities"*,
        Internal AD4 Paper, copies available from the authors on
        request.

[7]   Finch, W., Magill, S.A.N., Tyler, A.C.F, & Wilkinson, E.T.,
        "An Air Traffic Simulator for ATC Experiments", RSRE Memo
        3817, (in press), 1985.

[8]   Janda, A. (ed)., Proceedings of CHI'83, *"Human Factors in
        Computing Systems"*, Boston, 1983.

[9]   Leplat, J., & Bisseret, A., *"Analyse des processus de
        traitement de l'information chez le controleur de la
        navigation aerienne"*, Bulletin due Centre de'Etudes et de
        Recherches Psychotechniques, 14, 51-67,1965.

[10]   Leplat, J., & Hoc, J.M., *"Subsequent Verbalisation in the
        Study of Cognitive Processes"*, Ergonomics, 24, 743-755,
        1981.

[11]   Magill, S.A.N., & Ord, G., *"Air Traffic Management Concepts
        for the UK in the 1990's - a Review of Recent Work at RSRE"*,
        RSRE Memo 3714, 1984.

[12]   Neal, Alan S., & Simons, Roger M., *"Playback: A method for
        evaluating the usability of software and its documentation"*,
        HFC-48 Human Factors Center, IBM General Products Division,
        San Jose, California, September 1983.

[13]   Proceedings of Conference on *"Human Factors in Computer
        Systems"*, National Bureau of Standards, Gaithersburg, March
        1982.

[14]   Ratcliffe, S., *"Application of a Tape-line Height Measuring Device to Traffic near an Airway Intersection"*, RSRE Memo 3419, 1981.

[15]   Sperandio, J.C., *"Extensions to the Study of the Operational Memory of Air Traffic Controllers"*, RSRE Translation No 518, 1974.

[16]   Stamp, R.G., & Smith A.D.N., *"A Method for Estimating the Capacity of Air Traffic Sectors"*, DORA Research Paper 7402, October 1974.

[17]   Umbers, I.G., *"A Study of Control Skills in an Individual Task, and a Simulation, using the Verbal Protocol Technique"*, Ergonomics, 24(4), 275-293, 1981.

[18]   Weilinga, B.J., & Breuker, J.A., *"Interpretation of Verbal Data for Knowledge Acquisition"*, in ECAI'84: Advances in Artificial Intelligence", Elsevier, 1984.

[19]   Whitfield, D., *"A preliminary Study of the Air Traffic Controller's Picture"*. Journal of the Canadian Air Traffic Controllers' Association, ii(i), 19-28, 1979.

[20]   Whitfield, D., Ball, R.G., & Ord, G., *"Some Human Factors Aspects of Computer-Aiding Concepts for Air Traffic Controllers"*, Human Factors, 22(5), 569-580, 1980.

[21]   Whitfield, D., & Jackson, A., *"The Air Traffic Controller's 'Picture' as an example of a Mental Model"*. Proceedings of the IFAC/IFIP/IFORS/IEA "Conference on Analysis, Design and Evaluation of Man-Machine Systems", Baden-Baden, Fed. Rep. Germany, 1982.
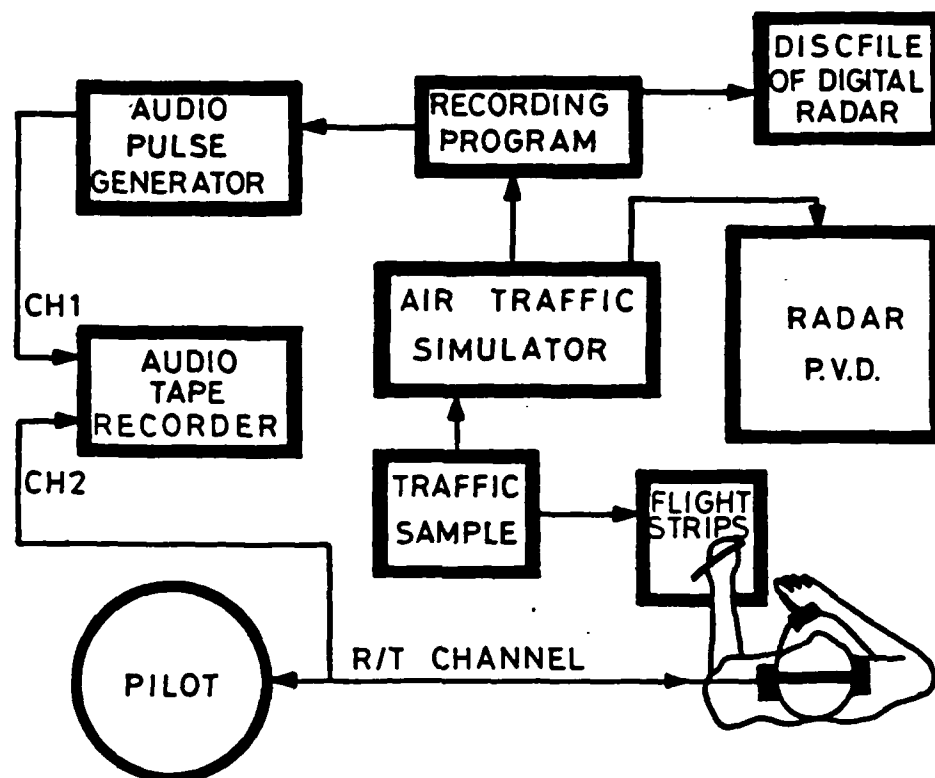
ACKNOWLEDGEMENTS

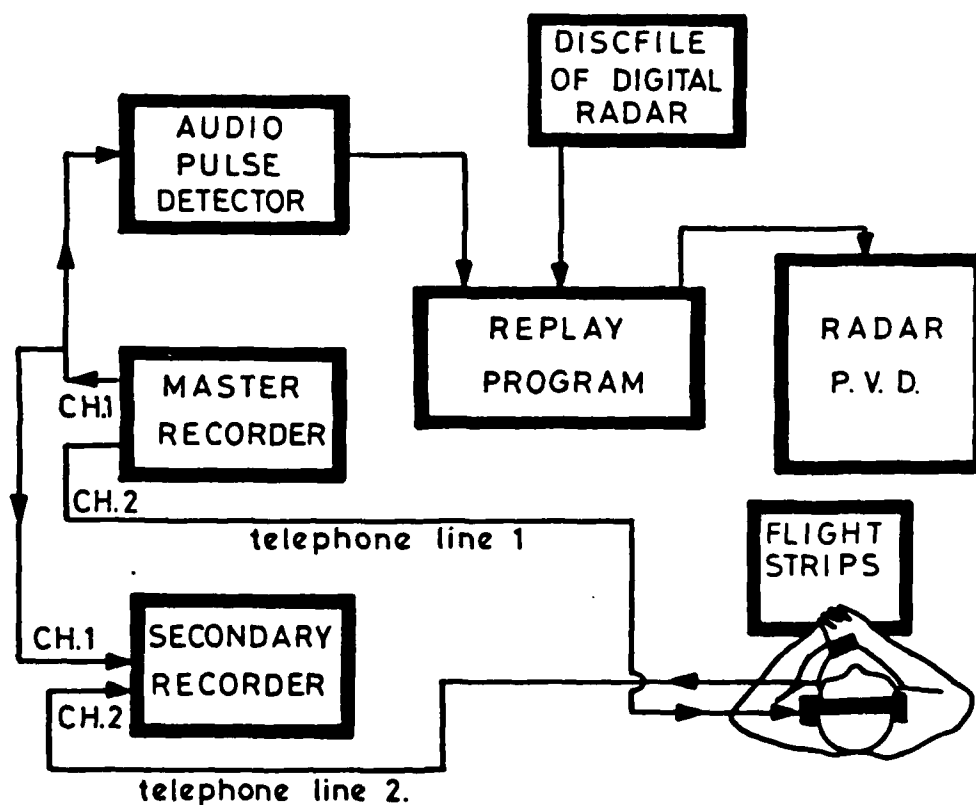FIG.1.   THE REPLAY SYSTEM IN RECORDING MODE

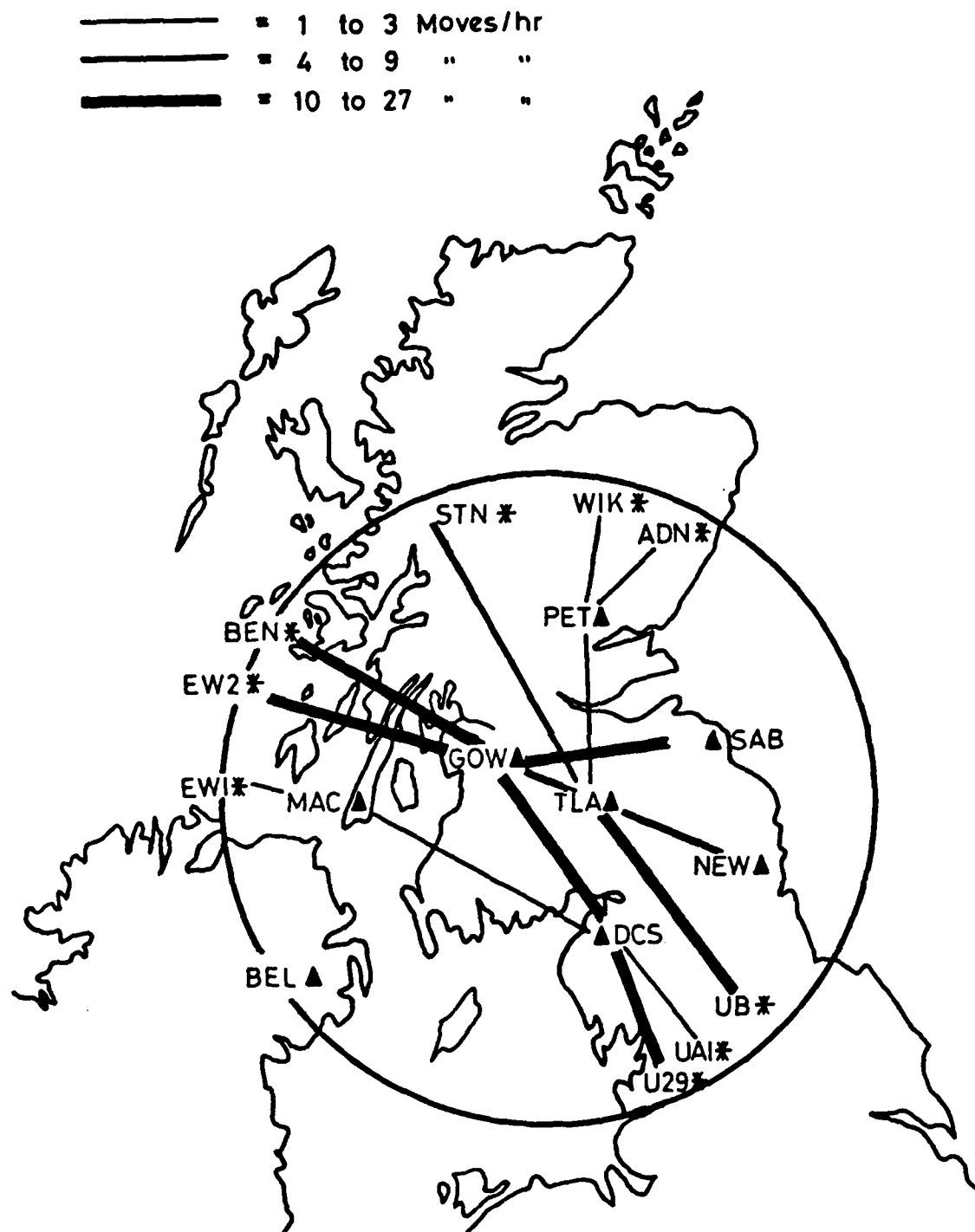FIG.2.   THE REPLAY SYSTEM IN REPLAY MODE

FIG. 3. ROUTE STRUCTURE EMPLOYED IN THE PILOT STUDIES
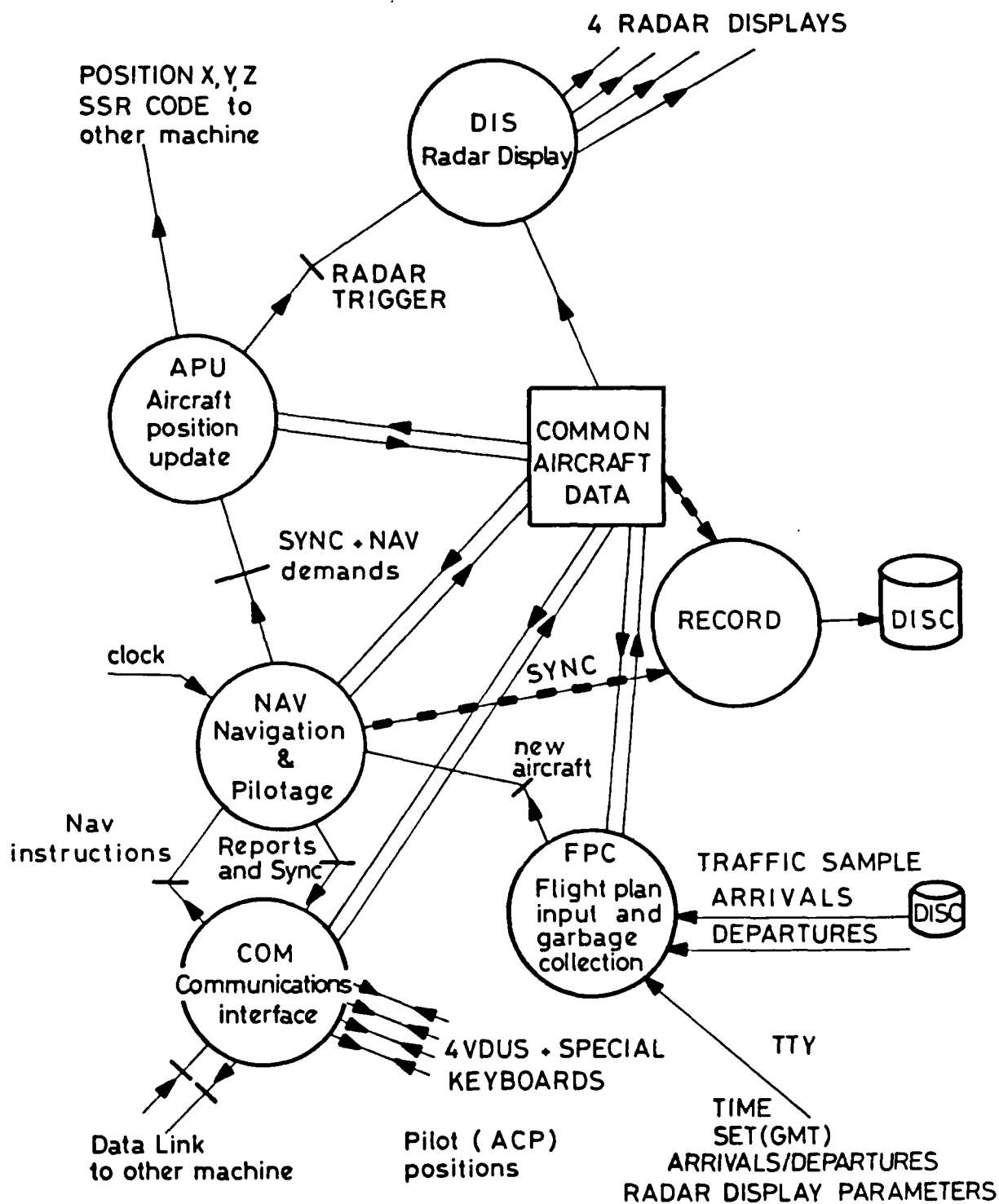(reproduced from RSRE MEMO. 3419)

FIG. 4. SIMULATOR SOFTWARE STRUCTURE EXPANDED TO INCLUDE
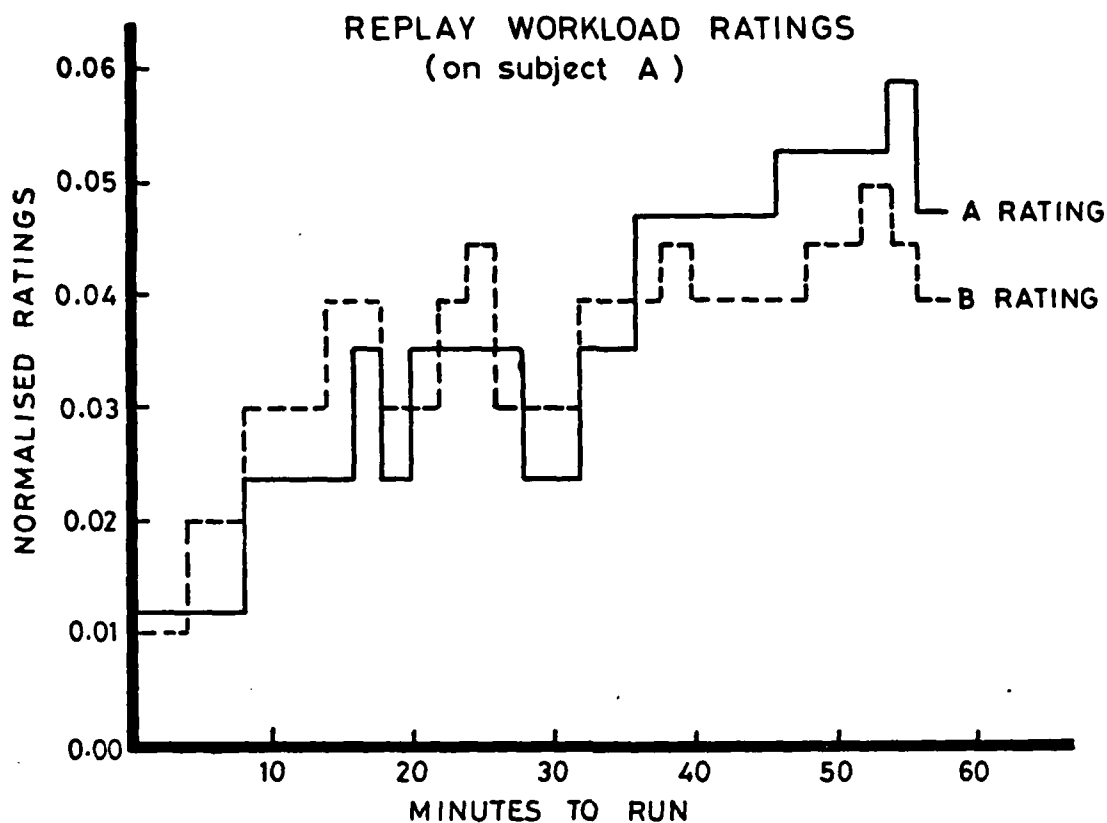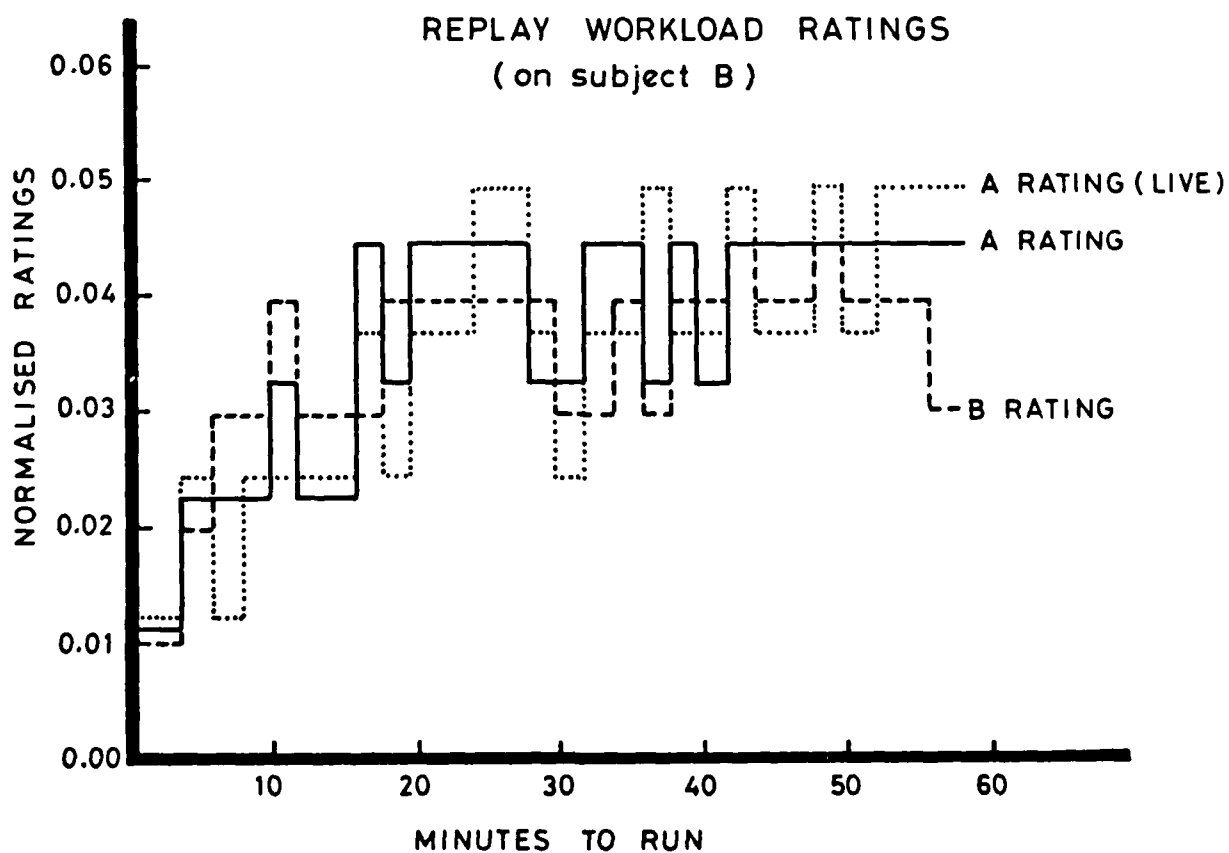REPLAY RECORDING (adapted from a figure in ref. 6.)

FIG. 5a.

FIG. 5b.   FIGURES 5a and 5b WORKLOAD RATING DATA FLOW PILOT STUDY 1.

COL 1 Time	COL 2 RT Transaction	COL 3 Replay Protocol	COL 4 Description	COL 5 Comments on Description
05.00 (030)	Over CONT Thank you CONT Springbok 2210 *turn right* *heading 160* (RT POOR)			
05.10 (031)	A/C Springbok 2210 roger Scottish, turning right from 150 to 160 degrees over.	That should bring him back towards the centre of track.	Should bring 2210 on a course roughly parallel to the opposing 2414 but displaced about 12 miles.	*It will keep him in the picture (not too far to edge of display) and also give earlier crossing on the Yugoslav (JG3842 on crossing course).*
05.20 (032)	A/C OM1186 *on your* *frequency at . . . .* cruising at FL350 U25 for EW.	The Iranian 3846 is in conflict with the Yugoslav . . . . they are about 60 miles distant at this time.		
05.30 (033)	1. over CONT 186 roger cleared present position, direct EW1. A/C OM1186 roger Scottish taking up heading now direct for EW1		Taken OM1186 completely clear of everything else	**Especially the AZ and avoids Dean's Cross.**

Figure 6   Example of verbal data using replay techniques from a simulation experiment.
 CONT = Controller, A/C = Aircraft, bracketted statements are explanatory insertions by experimenter.

DOCUMENT CONTROL SHEET

Overall security classification of sheet ....UNCLASSIFIED.. ....

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the box concerned must be marked to indicate the classification eg (R) (C) or (S))

1. DRIC Reference (if known)	2. Originator's Reference MEMORANDUM 3827	3. Agency Reference	4. Report Security Classification
5. Originator's Code (if known)	6. Originator (Corporate Author) Name and Location ROYAL SIGNALS AND RADAR ESTABLISHMENT		
5a. Sponsoring Agency's Code (if known)	6a. Sponsoring Agency (Contract Authority) Name and Location CIVIL AVIATION AUTHORITY C L HYNAM CR(8) M J ASBURY RD1		

7. Title
THE REPLAY TECHNIQUE: THE CONCEPT, INITIAL EXPERIENCE AND PROPOSED DEVELOPMENTS

7a. Title in Foreign Language (in the case of translations)

7b. Presented at (for conference papers)   Title, place and date of conference

8. Author 1 Surname, Initials JACKSON A	9(a) Author 2 ONSLOW G J	9(b) Authors 3,4...	10. Date	pp. ref.
11. Contract Number	12. Period	13. Project	14. Other Reference	

15. Distribution statement
        UNLIMITED

Descriptors (or keywords)

continue on separate piece of paper

Abstract

    A system is described for capturing activity at the Human-Computer Interface in complex systems. This system, implemented in a Simulated Air Traffic Control Environment, allows for the subsequent replay of all task related, audio and visual inputs to the controller, synchronised with a record of his own manual and spoken system inputs. Potential applications for such a facility are discussed and three pilot studies considering the tool's usefulness in examining operator workload and assisting in system development are reported. A revised version of the facility, for use in an enhanced simulation environment is also described.

S80/48

END
DTIC
7-86